UC Berkeley
UC Berkeley Previously Published Works

Title
Revealing ferroelectric switching character using deep recurrent neural networks

Permalink

Journal

ISSN

Authors
Agar, Joshua C
Naul, Brett
Pandya, Shishir
et al.

Publication Date

DOI

Peer reviewed

ARTICLE

Revealing ferroelectric switching character using deep recurrent neural networks

Joshua C. Agar[1,2,3]*, Brett Naul [4], Shishir Pandya[1], Stefan van der Walt [5], Joshua Maher[1], Yao Ren[6], Long-Qing Chen[7], Sergei V. Kalinin [8], Rama K. Vasudevan [8], Ye Cao[6], Joshua S. Bloom[4] & Lane W. Martin [1,2]*

The ability to manipulate domains underpins function in applications of ferroelectrics. While there have been demonstrations of controlled nanoscale manipulation of domain structures to drive emergent properties, such approaches lack an internal feedback loop required for automatic manipulation. Here, using a deep sequence-to-sequence autoencoder we auto-mate the extraction of latent features of nanoscale ferroelectric switching from piezo-response force spectroscopy of tensile-strained $PbZr_{0.2}Ti_{0.8}O_3$ with a hierarchical domain structure. We identify characteristic behavior in the piezoresponse and cantilever resonance hysteresis loops, which allows for the classification and quantification of nanoscale-switching mechanisms. Specifically, we identify elastic hardening events which are associated with the nucleation and growth of charged domain walls. This work demonstrates the efficacy of unsupervised neural networks in learning features of a material's physical response from nanoscale multichannel hyperspectral imagery and provides new capabilities in leveraging in operando spectroscopies that could enable the automated manipulation of nanoscale structures in materials.

[1] Department of Materials Science and Engineering, University of California, Berkeley, Berkeley, CA 94720, USA. [2] Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. [3] Department of Materials Science and Engineering, Lehigh University, Bethlehem, PA 18015, USA. [4] Department of Astronomy, University of California, Berkeley, Berkeley, CA 94720, USA. [5] Berkeley Institute of Data Science, University of California, Berkeley, Berkeley, CA 94720, USA. [6] Department of Materials Science and Engineering, The University of Texas at Arlington, Arlington, TX 76019, USA. [7] Department of Materials Science and Engineering, Pennsylvania State University, University Park, PA 16802-5006, USA. [8] Center for Nanophase Materials Sciences, Oak Ridge National Laboratory, Oak Ridge, TN 37830, USA. *email: joshua.agar@lehigh.edu; lwmartin@berkeley.edu

The ability to create and manipulate domain structures in ferroelectrics allows for the control of the phase and polarization orientation and the local and macroscale susceptibilities (e.g., electrical, thermal, mechanical, optical), thus providing a foundation for next-generation devices[1–3]. Early efforts in this regard have focused on deterministically creating desired domain structures. In, for example, tetragonal $PbZr_{1-x}Ti_xO_3$, controlling the elastic boundary conditions has provided access to domains spanning simple monodomain to complex hierarchical domain structures[4–6]. As the field has advanced, ingenious routes, including compositional gradients[7,8], superlattice structures[9–11], orientation control[12,13], and engineered octahedral rotations[14], have been leveraged to control domain structures.

The majority of this work, however, has focused on the static creation of desired domain structures[6,15–18] or functional domain walls[19,20] and thus lacks an internal self-regulating feedback loop required for automatic operation in functional devices. To deterministically manipulate ferroelectric domain structures requires the ability to measure while in operation (i.e., in operando) and automatically identify a number of features (e.g., polarization orientation, switching pathways, domain-wall geometry). Developments in multimodal spectroscopy now allow the acquisition of data at both the appropriate time- and length scales required to glean such information from ferroelectric materials using techniques such as transmission electron microscopy[21,22], scanning-probe microscopy[23,24], diffraction studies[25,26], etc.[27] The challenge, however, is that downstream analytical approaches which project data into a human-interpretable form remain underdeveloped and ill equipped for the complexity and magnitude of the data that can now be readily produced. In turn, despite the extensive amount of expensive experiments conducted, only an infinitesimally small fraction of the data collected is translated into knowledge.

Solving this challenge requires looking beyond the borders of nanoscience to fields such as social analytics[28,29], natural-language processing[30,31], and sentiment analysis[32,33], where computational roadblocks are pervasive. For decades, standard practice was to develop machine-learning algorithms to create mathematical abstractions of the data based on characteristics of preconceived importance. Recently, the availability of massive datasets and specifically designed hardware has enabled features once designed by human experts to be extracted using brute-force computation. These representation learning tools generally rely on building architectures of simple non-linear mathematical functions which are optimized to relate the raw data to some information or label[34–36]. These so-called deep-learning-neural-network-based approaches have set new benchmarks for many common machine-learning tasks including image[37] and speech recognition[38], language translation[39], and identification of human intention[32,33]. While these deep-learning approaches have begun to make meaningful inroads in, for example, genomics[40], high-energy physics[41], and astronomy[42], they have yet to be sufficiently embraced in experimental nanoscience[43–53].

Here, we develop a sequence-to-sequence neural network to extract inference from band-excitation piezoresponse spectroscopy (BEPS). To test our approach, we conducted BEPS on tensile-strained $PbZr_{0.2}Ti_{0.8}O_3$ thin films wherein strain drives the formation of a hierarchical c/a and a_1/a_2 domain structure. We develop and train a deep-learning-neural-network-based sparse autoencoder on piezoresponse hysteresis loops to demonstrate parity with conventional empirical-analysis approaches. We then apply this approach to extract insight from the resonance response which has a form too complex to be properly analyzed using techniques common in experimental materials science. Using the information "learned", we identify geometrically driven differences in the switching mechanism which are related to

charged-domain-wall nucleation and growth during ferroelastic switching. This insight could not have been extracted using machine-learning approaches that have been previously applied to materials spectroscopy and provides unprecedented information about the nature of the specific domain-structure geometries that should be explored to enhance local and macroscale susceptibilities. Furthermore, the ability to automate the extraction of inference regarding ferroelectric-switching mechanisms from multichannel nanoscale spectroscopy provides the first step (i.e., machine-learned discrimination) that could be used to design real-time control systems capable of creation and verification of interconversion of functional domain structures and interfaces. The developed approach is extensible to other forms of multidimensional, hyperspectral images (wherein there is a spectra at each pixel) which are commonly acquired in experiments such as time-of-flight secondary-ion mass spectrometry[54,55], scanning Raman[56], electron energy loss spectroscopy[57,58], etc. To promote the utilization of this approach, we provide open access to all data and codes in the form of a Jupyter notebook[59] (Supplementary Note 1). Ultimately, this work represents an example of how unsupervised deep learning can highlight features relating to ferroelectric physics overlooked by human-designed-machine-learning algorithms, and how such approaches can be adapted to analyze hyperspectral data more broadly.

Results

Synthesis of $PbZr_{0.2}Ti_{0.8}O_3$ thin films with hierarchical domain structures. We synthesized $PbZr_{0.2}Ti_{0.8}O_3/Ba_{0.5}Sr_{0.5}RuO_3/NdScO_3$ (110) heterostructures using pulsed-laser deposition (Methods). The resulting films have a hierarchical domain structure with a sawtooth topography on two length scales (Fig. 1a–e), as the result of primarily out-of-plane polarized $c/a/c/a$ [with enhanced out-of-plane (Fig. 1b) and suppressed in-plane (Fig. 1c) piezoresponse] and fully in-plane polarized $a_1/a_2/a_1/a_2$ [with suppressed out-of-plane and enhanced in-plane piezoresponse] domain bands. This hierarchical domain structure emerges due to the tensile strain which drives the c/a and a_1/a_2 domain variants to be nearly energetically degenerate (Supplementary Note 2)[6].

Band-excitation piezoresponse force microscopy. To characterize the nanoscale-switching processes, we conducted BEPS (Methods and Supplementary Note 3, Supplementary Figs. 2 and 3). Briefly, BEPS measures the piezoresponse in remanence (across a band of frequencies near the cantilever resonance) following a perturbation from a bipolar-triangular switching waveform designed to fully switch the material. Following fitting, the raw data from this experiment at every pixel (x, y size = 60, 60) measures the amplitude (A), phase (ϕ), resonance frequency (ω), and quality factor (Q) of the cantilever resonance, which are qualitative measures of the piezoresponse, polarization direction, stiffness, and dampening, respectively, at various voltages (V length = 96). At the most basic level, visualization of the switching processes can be achieved by creating movies from the images of the signals throughout the switching process or by plotting the response curves at a specific location or within a predefined area (Supplementary Movie 1). Additionally, it is common to compute at each tip position a piezoresponse loop ($A \cos \phi$) which can then be fit to a 15-parameter empirical function. While such approaches are capable of visualizing large differences in the piezoresponse they provide only limited information into subtle differences of the response, which contains important insight (Supplementary Note 4, Supplementary Fig. 4). While in the raw form the data might occupy an N-dimensional space, the information of physical significance lies on
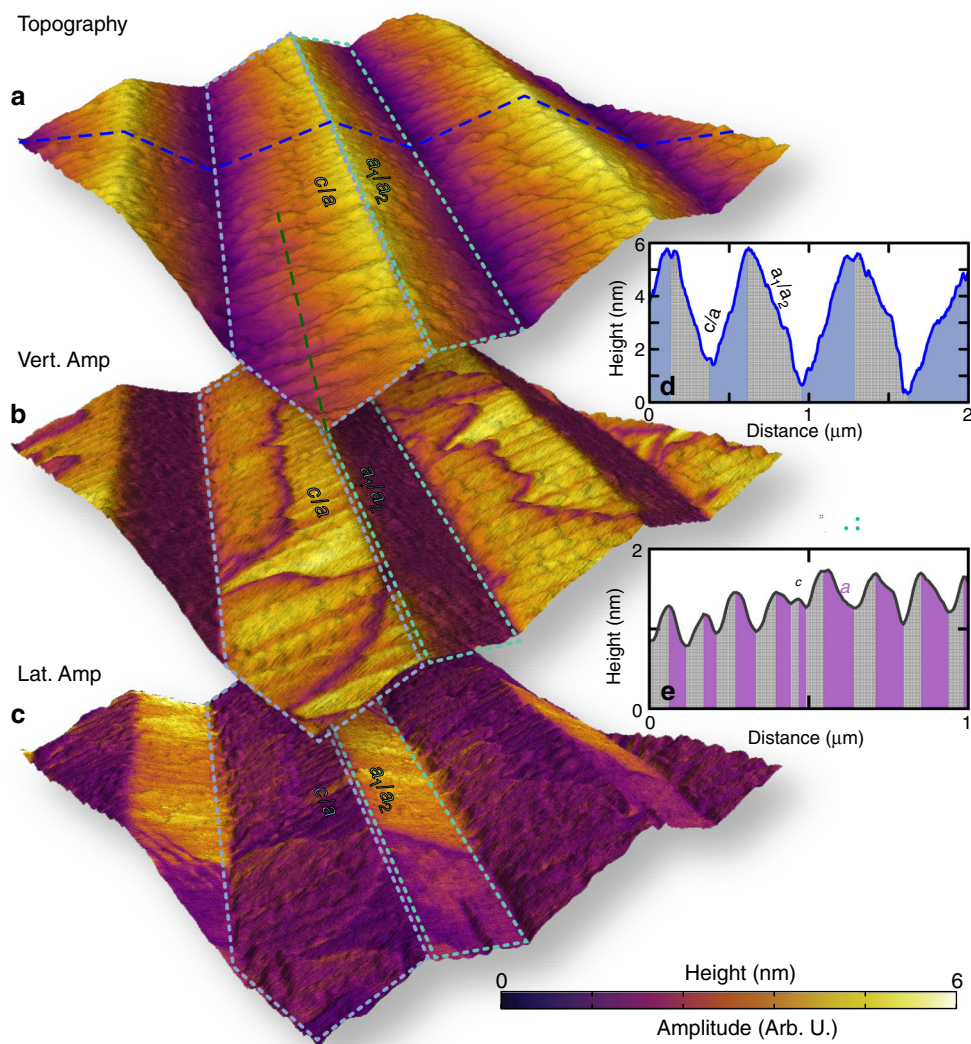
Fig. 1 Surface topography of $PbZr_{0.2}Ti_{0.8}O_3$ with hierarchical domain structures. **a** Three-dimensional tapping mode topography superimposed with topographic information presented in nm as indicated. **b**, **c** Three-dimensional tapping mode topography superimposed with the **b** vertical and **c** lateral piezoresponse amplitude. Color presented in arbitrary units defined by the color scale in the inset. Topographic line traces showing the surface topography across the **d** c/a-a_1/a_2 bands (indicated by the dark-blue-dashed line in **a**) and **e** c/a bands (indicated by the dark-green-dashed line in **a**)

a data manifold with a much lower dimensionality; however, we have no means to predict the manifold.

Recognizing these limitations, statistical approaches of machine learning have been applied to predict the data manifold, thus allowing more insight to be extracted from BEPS data. To demonstrate the necessity of the proposed deep-learning approach, we have conducted careful analysis using linear decomposition algorithms, including principal-component analysis and non-negative-matrix factorization, and a variety of clustering algorithms (Supplementary Note 5–7, Supplementary Figs. 5–7). All told, while these methods are able to identify the most significant differences in the response, they have minimal sensitivity to quantify subtle, yet physically significant, features. This is at least in part due to the decoupling of the voltage (i.e., the time or temporal) dependence of the spectra which are viewed as independent samples—meaning that these algorithms exclude the history of when each data point was collected.

Inference from long–short-term memory recurrent autoencoder. Thus, what is required is to develop an approach that

considers the temporal dependence inherent in the data. To do this, we developed a sequence-to-sequence deep learning neural network based on a long–short-term memory (LSTM) recurrent neural network (RNN) autoencoder (henceforth called the autoencoder) which acts as a feature extractor to derive inference from BEPS spectra. LSTM neurons (described in Supplementary Note 8, Supplementary Fig. 8) were chosen due to their success in natural language processing (which is analogous in data structure to spectra) wherein order of words (measurements) is important[60]. The autoencoder architecture consists of an encoder, which takes as an input a spectra and outputs a feature vector, and a decoder, which takes this feature vector and returns the input spectra (Fig. 2). By minimizing the mean-squared-reconstruction error of the input spectra the autoencoder "learns" a universal identity function. While building an arbitrarily complex identity function (where the large model capacity assures overfitting) is a fruitless task, building an identity function whose capacity is limited or strongly regularized can produce a generalizable (i.e., with a limited number of characteristic variables) function capable of broader inference.
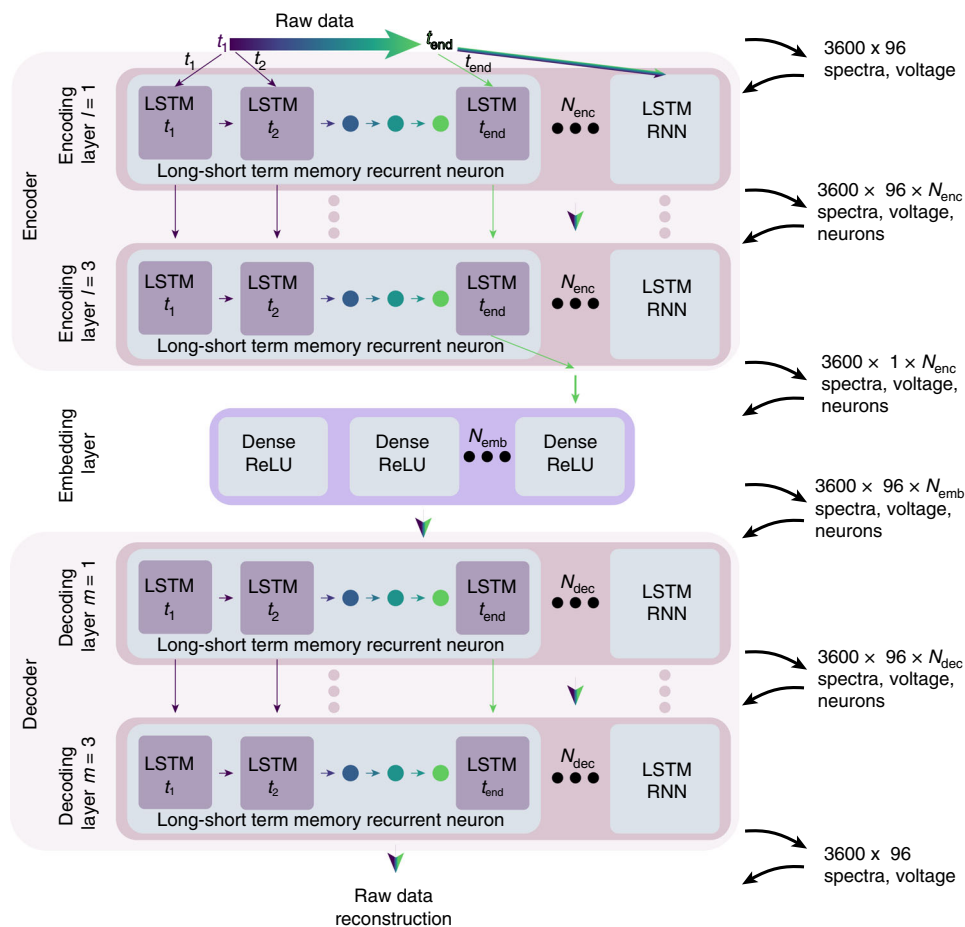
Fig. 2 Drawing of sparse long–short-term memory autoencoder. Diagram shows the three components of the neural network the encoder, embedding layer, and decoder. Within each of these sections the dimensionality and size of the input and outputs are indicated on the right. Diagram shows how temporal data (represented as linear color-changing arrows) is consider through the inclusion of recurrent long–short-term memory neurons. Solid colored arrows indicate just a single time step is passed, where arrows with gradients imply the passing of temporally dependent vectors. In the figure l is the number of encoding layers, m is the number of decoding layers, N_{enc} is the number of neurons in the encoding layer, N_{emb} is the number of neurons in the embedding layer, and N_{dec} is the number of neurons in the decoding layer

More explicitly, the first encoding layer accepts a time series as an input which, in this case, is the response from the 96 sequential voltage steps. Each of the encoding layers is composed of LSTM neurons ($N_{enc} = 256$)[61] with an internal cell structure that enables the retention of long- and short-term temporal dependencies. During training, each layer outputs an abstract representation of the data for each input spectra. To build more descriptive data abstractions, it is typical to stack encoding layers ($l = 3$) forming so-called deep neural networks. To minimize overfitting or memorization, dropout (d = 20%)[62] (which minimizes co-adaptation by randomly removing some of the network connections) was applied. The output from the encoder is then passed to an embedding layer consisting of dense neurons ($N_{emb} = 16$)[63]. This layer creates a low-dimensional representation of different characteristic responses. To make this representation more interpretable, it is beneficial to impose sparsity [i.e., minimize the number of non-zero activations (outputs)]. To do this, two synergistic approaches were applied: First, we restrict the outputs to non-negative values by selecting a rectified-linear activation function (ReLu) ($f(x) = \max(0, x)$). Secondly, we add strong l_1 regularization which adds an additional contribution to the loss function proportional to the sum of the weights $(\lambda \sum_i |W_i|)$, thus only those activations which significantly improve the model's accuracy are non-zero (this can be visualized

in Supplementary Movies 2 and 3). The feature vector is then passed to the decoder which is structured in an identical fashion to the encoder [decoding layers ($m = 3$), each with 128 LSTM neurons ($N_{dec} = 128$)]. The decoder takes the feature vector and transforms it back into the original spectra, such that the network can be optimized to minimize the loss function composed of the mean-squared error and the l_1 regularization $[(\text{loss} = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 + \lambda \sum_i |W_i|)$, Supplementary Note 9, Supplementary Fig. 9].

We begin by training the autoencoder to analyze the piezo-response hysteresis loops extracted from BEPS (Methods). Since every spectra analyzed by the autoencoder has a known pixel position it is possible to visualize the "learned" information by computing the output of the low-dimensional layer to form a real space feature map. While this layer would permit an independent feature map for each neuron, the addition of sparsity results in most features having a null value. Here, we show two of the most physically meaningful features (Fig. 3a, b), where the intensity of the maps represents the degree of a "learned" characteristic response form. To aid in the visualization of this information we have provided a line trace of the average activation superimposed on the average topography (Fig. 3c, d). The first feature map (Fig. 3a) shows increased activation within the c/a bands which is maximized on the peak side of the topographical features
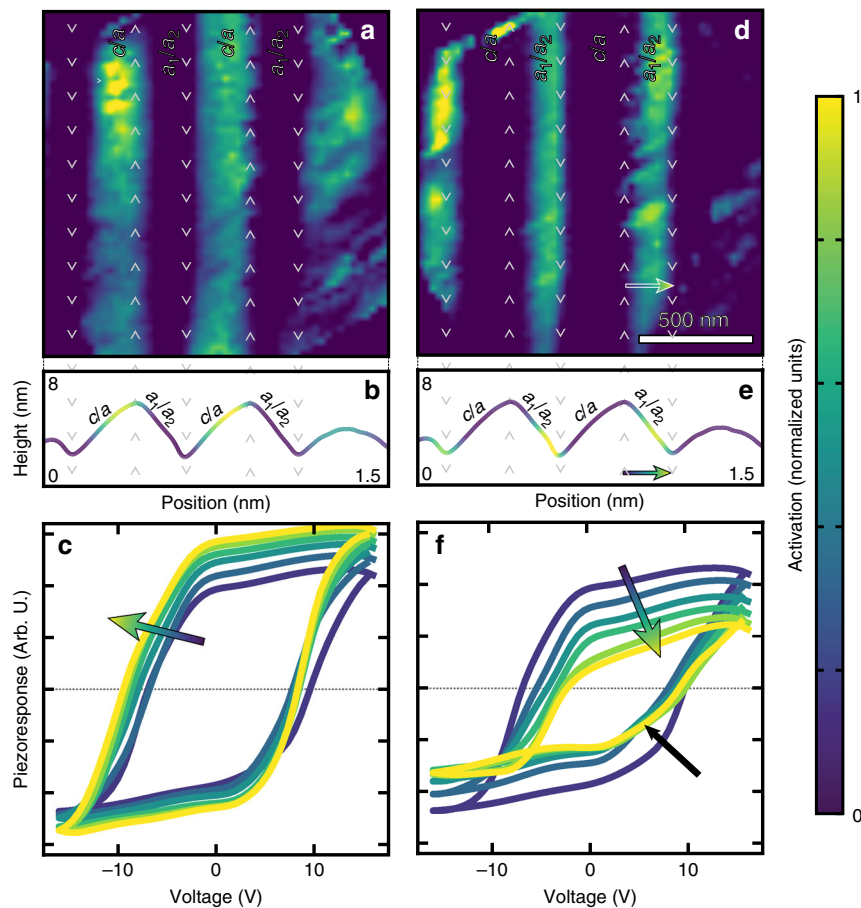
Fig. 3 Features learned from low-dimensional layer of the piezoresponse autoencoder. **a**, **d** Feature maps extracted from low-dimensional layer of autoencoder trained on piezoresponse hysteresis loops. Color indicates the magnitude of the latent feature or the activation observed in each spectra at a given pixel position. Activation is mapped in normalized units as shown in colorbar. **b**, **e** Average activation across the domain bands superimposed onto the average topography. **c**, **f** Neural network generated piezoelectric hysteresis loops as the magnitude of the activation is increased. In all figures the color of the curves/images reflect the normalized activation from the low-dimensional representation at that location or from the generated response curve

(Fig. 3c). To visualize what this activation is encoding, we use the decoder of the autoencoder as a generator, allowing us to manually manipulate the activation to see how this neuron alters the piezoresponse (Supplementary Movie 4). From the generated piezoresponse hysteresis loops, we observe an increase in the magnitude and the squareness of the loops as we increase the activation of this neuron, as would be expected for regions with increasing c-like character (Fig. 3e). This suggests that the neuron has "learned" the response associated with c domains and, since the map has different magnitudes of response, it provides a way to quantify the c-like switching character spatially. In total, this reveals that the autoencoder is capable of deducing physically interpretable inference from the data in an unsupervised fashion. Moving on to the second feature map (Fig. 3b), we notice a gradient in the activation which is maximized within the a_1/a_2 bands near the valley boundary which decreases in magnitude as we transition towards the peak boundary (Fig. 3d). Visualizing the loops as we increase this activation reveals a decrease in the magnitude and the emergence of an intermediate step in the piezoresponse loop (black arrow, Fig. 3f). This intermediate step is related to a two-step, three-state ($c \rightarrow a \rightarrow c$) ferroelastic switching process[6].

It is important to reemphasize that this detailed spatial variance was not apparent using conventional machine-learning approaches. All told, this reveals that the autoencoder is capable of "learning" a complex identity function where each neuron controls a physically meaningful characteristic of the

piezoresponse. Not only is the autoencoder readily able to differentiate regions of varied response which correlate to different domain structure variants, but it is also able to quantify the relative response character (i.e., how "x-like" the response is) which reveals additional complexity. This deviates from traditional approaches which generally provide only qualitative classification of behavior.

Having proven the capabilities of this approach, we applied a similar methodology to interpret the cantilever-sample-contact resonance (henceforth the resonance response) wherein the form of the response has increased complexity which complicates statistical analysis. We show three of the most physically meaningful non-zero components (Fig. 4), obtained following training (Methods). To use the extracted insight to understand the physical mechanisms of response it is important to identify the statistical distribution of the key response characteristics "learned". The first selected feature map (Fig. 4a) shows increased activation within the c/a bands which is maximized near the valley boundary (Fig. 4b). If we again use this autoencoder as a generator (Supplementary Movie 5), we observe that the resonance has a classic butterfly-shaped loop, however, as this activation increases there is a gradual increase in the resonance frequency as we approach the valley boundary (Fig. 4c). If we had followed the traditional approach of merely studying how the piezoresponse hysteresis loops vary as we move from the peak to the valley boundary (Fig.4d), we would have observed essentially no change, thus leaving uncovered new physical effects.
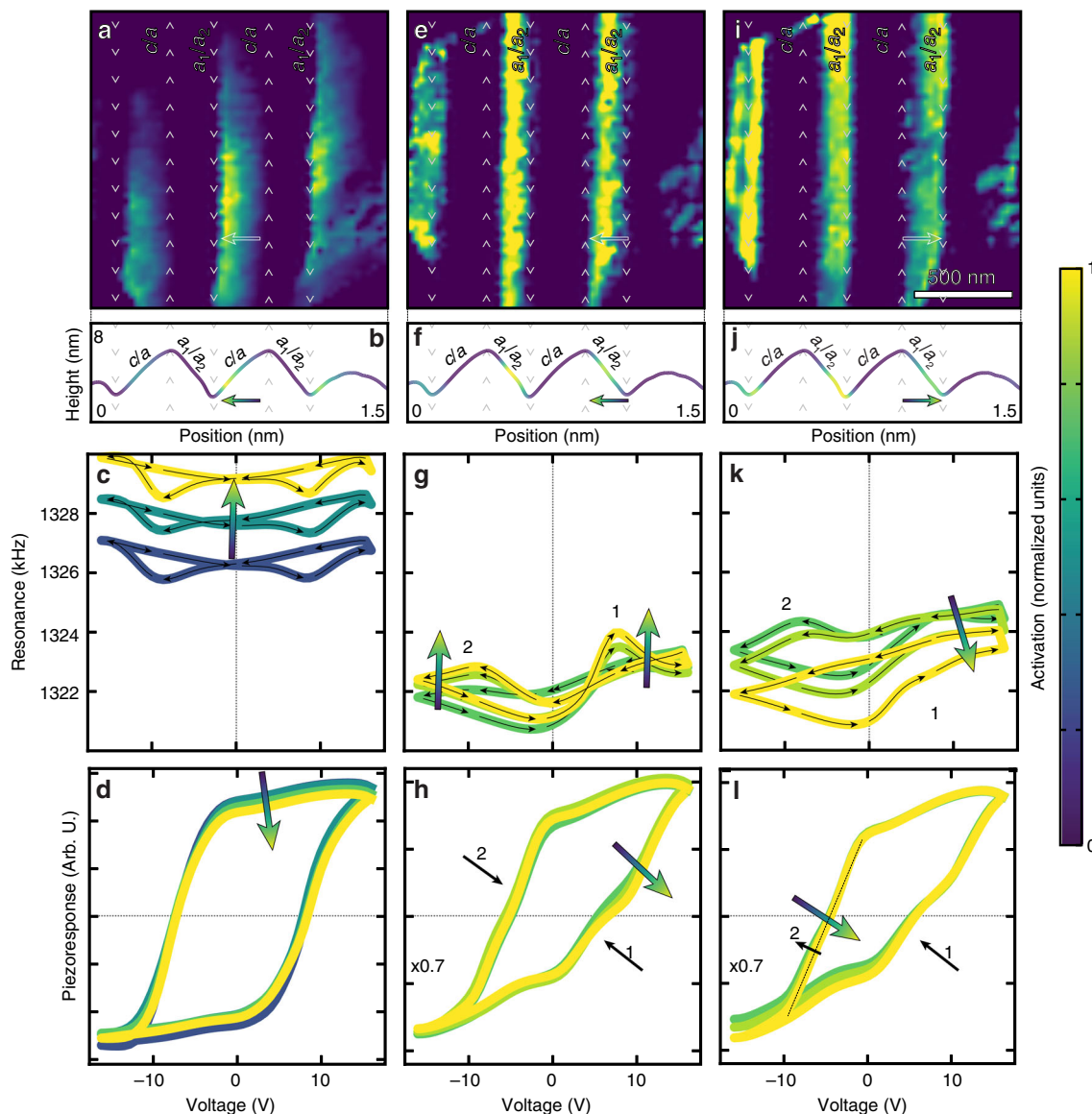
5

Fig. 4 Features learned from low-dimensional layer of the resonance response autoencoder. **a**, **e**, **i** Feature maps extracted from low-dimensional layer of autoencoder trained on the resonance hysteresis loops. Color indicates the magnitude of the latent feature or the activation observed in each spectra at a given pixel position. Activation is mapped in normalized units as shown in colorbar. **b**, **f**, **j** Average activation across the domain bands superimposed onto the average topography. Neural network generated **c**, **g**, **k** resonance hysteresis loops and **d**, **h**, **l** piezoresponse hysteresis loops. In all figures the color of the curves/images reflect the normalized activation from the low-dimensional layer at that location or from the generated response curve. Numbers in figures represent observations of ferroelectric or ferroelastic switching events

Moving on to the second feature map (Fig. 4e), we observe a gradient in the activation, which is maximized near the middle of the a_1/a_2 band and tends towards zero at the valley boundary (Fig. 4f). The generated resonance loops reveal a non-traditional shape wherein upon application of bias the material undergoes elastic hardening before eventually softening when switching under both positive and negative bias (marked 1 and 2, respectively, Fig. 4g). This type of resonance behavior during switching has been related to a two-step, three-state ($c \rightarrow a \rightarrow c$) ferroelastic switching mechanism[6]. By computing the piezo-electric hysteresis loops, we identify that these loops have an intermediate step which plateaus at near-zero piezoresponse when switching under both positive and negative bias (1 and 2, respectively; Fig. 4h). The generated piezoresponse curves reveal that as we decrease the magnitude of this activation (i.e., traverse the a_1/a_2 band from the mid-point to the valley) we observe an

increase in the prevalence (magnitude of the concavity) of this intermediate state when switching under positive bias; however, such a change is not evident when switching under negative bias.

Finally, the third feature map (Fig. 4i) once again shows increased activation within the a_1/a_2 bands; however, the gradient in activation goes in the opposite direction from the previous map (Fig. 4e), in that it is maximized near the valley boundary and decreases as we approach the middle of the band (Fig. 4j). Upon generating the resonance (Fig. 4k) and piezoresponse (Fig. 4l) loops, we observe an asymmetric resonance switching behavior, wherein near the valley boundary (i.e., high activation) hardening only occurs when switching under positive bias (marked 1). As we move towards the peak boundary (i.e., low activation), however, hardening during switching occurs under both positive and negative bias (green curves, Fig. 4h). An analogous trend in the piezoresponse concavity is observed (Fig. 4l), where the

intermediate concavity is only observed when switching under positive bias when near the valley boundary (marked 1); yet, is observed under both positive and negative bias when near the peak boundary (green curve, marked 2).

Phase-field simulations of ferroelectric switching. To understand the physical significance of the features identified by the autoencoder requires consideration of the domain structure, geometry, and switching processes. To guide our interpretation, we conducted phase-field switching studies (of a model film with a c/a domain structure) under a simulated tip bias (Methods, Supplementary Note 10, Supplementary Fig. 10). From these phase-field studies, we can observe the polarization as well as the local energetics during the switching process. Prior to the switching studies, the film exists in the up-poled state where the film has nearly uniform electrostatic energy.

Starting with the first "learned" feature, which is most pronounced when the tip is within the c domains near the valley boundary, we observe a square piezoelectric hysteresis loop (Fig. 5a) and an increase in the resonance frequency of the cantilever (i.e., an increase in the elastic modulus, Fig. 5b). From the initial state, phase-field simulations reveal that switching at this position results in the nucleation of a down-poled domain (Fig. 5c, top, Supplementary Movie 6). From the electrostatic energy, we observe a region of increased energy below the newly nucleated domain (Fig. 5c, arrow) as a result of the head-to-head charged domain wall which exists between the nucleated domain and the a domain. This electrostatic repulsion, from the growing charged domain wall, increases the local modulus of the material, which manifests as an increase in the cantilever resonance frequency. When applying negative bias to the tip, the bias reinforces the as-poled domain structure resulting in an unremarkable change in both the domain structure and local electrostatic energy (Fig. 5d).

Shifting our attention to the second "learned" feature, which is most pronounced at the peak of the a domains, where we observe intermediate concavities in the piezoresponse hysteresis loop (Fig. 5e) and elastic hardening when switching under positive and negative bias (Fig. 5f). Phase-field simulations reveal that switching at this location, under positive bias, results in the nucleation of a down-poled domain (Fig. 5g, top, Supplementary Movie 7) with a head-to-head positively charged domain wall. Turning our attention to the electrostatic energy we observe, as expected, a significant increase in the electrostatic energy near this charged domain wall (Fig. 5g, bottom, and indicated by the arrow). Thus, as the tip bias is increased, the newly nucleated domain grows laterally along the in-plane [100] and [010] resulting in an increase in the charged-domain-wall area. This, in turn, increases the stiffness, which manifests as an elastic hardening step. This hardening process continues until the finite volume probed by the tip is purely c-like, upon which application of further bias results in softening to saturation. When applying negative bias to the tip at this location, we observe the nucleation of an up-poled domain within the a domain (Fig. 5h, top). This up-poled domain must form a tail-to-tail negatively charged domain wall. In turn, the presence of the charged domain wall results in an increase in electrostatic energy (arrow, Fig. 5h, bottom) near this charged domain wall, which results in elastic hardening and softening to saturation with an identical mechanism as described for the behavior when switching under positive bias.

Finally, focusing on the third "learned" feature, which is most pronounced within the a domain near the valley, we observed intermediate concavities in the piezoresponse hysteresis loop (Fig. 5i) and elastic hardening when switching only under positive bias (Fig. 5j). Looking at the phase-field simulations under positive tip bias (Fig. 5k, Supplementary Movie 8), we observe the nucleation of a down-poled domain which is identical in form to that observed near the peak boundary within the a domain. As expected, the positive-bias branch of the piezoresponse loop and the resonance response have a similar form to the switching observed near the a-domain boundary, wherein the formation of a charged domain wall (Fig. 5k, bottom, arrow) results in an intermediate step in the piezoresponse loop and elastic hardening. When switching under negative bias, phase-field simulations reveal a different switching mechanism, wherein application of bias results in the expansion of the up-poled domain into the a domain, however, due to the geometry the domain is nominally uncharged (Fig. 5l). As a result, there is no evidence of either an intermediate concavity in the piezoresponse hysteresis loop nor a hardening step. All told, by interpreting the features learned by the autoencoder in the context of the domain structure and switching process we are able to identify features in both the piezoresponse hysteresis loops and resonance response related to the formation of charged domain walls during the switching process, which were not identified using conventional and linear machine learning analysis approaches.

Discussion

In summary, we demonstrate how a deep sequence-to-sequence, LSTM autoencoder can be used to "learn" characteristic mechanisms of response from multichannel hyperspectral BEPS which are overlooked by classical machine-learning approaches. We train a sparse LSTM autoencoder capable of identifying characteristic materials responses and quantify their relative significance with spatial resolution. The increased aptitude of this methodology stems from the inclusion of sequential dependence inherent in the data. Specifically, using this approach we experimentally measure and visualize changes in the elastic modulus associated with nucleation and growth of charged domain walls which form as a result of the hierarchal domain geometry in tensile-strained $PbZr_{0.2}Ti_{0.8}O_3$ thin films. This new capability provides a process to quantify subtle differences in switching mechanisms, providing a route towards the fully-automated and feedback-controlled nanoscale manipulation of domain structures and domain-wall geometries. Furthermore, this work establishes a new methodology capable of learning mechanisms of response from multichannel hyperspectral imagery which could be used for a range of existing and emerging hyperspectral imaging modalities in nanoscience.

More broadly, this work demonstrates a pathway towards the use of statistical approaches which take seemingly inconsequential data and, through mathematical transformations, creates observations from which meaning can be deduced. While such approaches have been used, most of these approaches have neglected the physical structure (e.g., symmetry, time, etc.) which provides a foundational basis for the data and the information which it contains. We emphasize how the sequential dependence of experimental data can be incorporated in such models, analogous concepts could be applied to include spatial position, symmetry, and both long- and short-range order, increasing their aptitude in inducing meaningful inference regarding materials systems and processes. This work paves the way for spectroscopic techniques wherein the conventional scientific methods of designing targeted experiments aimed at a specific hypothesis are supplanted by approaches which collect all seemingly relevant data, which can then be used to identify a hypothesis for empirical testing. Furthermore, this ability to automate the formation of fingerprints of physical processes from multichannel spectroscopies provides increased capabilities towards the
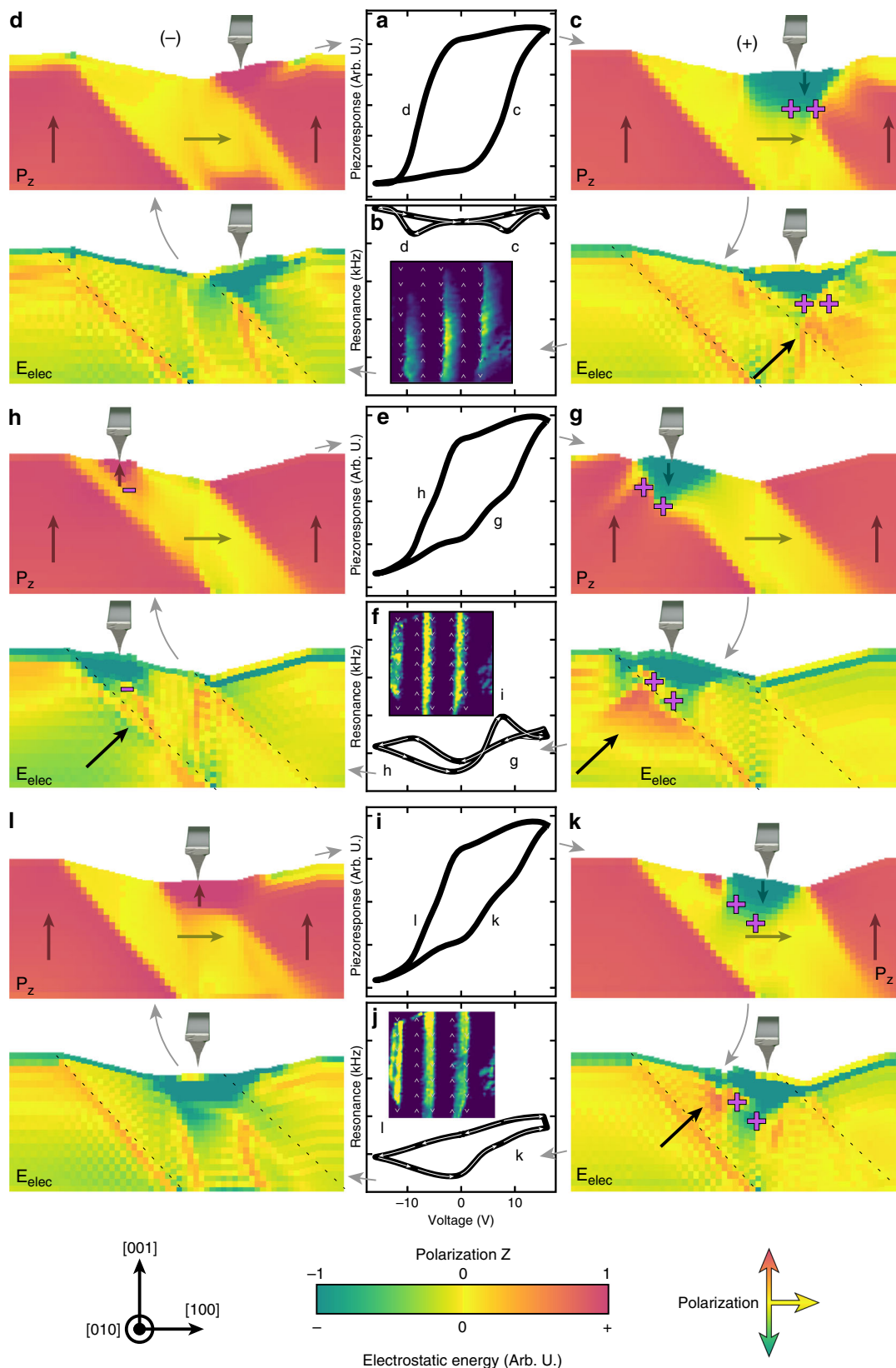
Fig. 5 Interpretation of learned features based on phase-field simulations. **a**, **e**, **i** Generated piezoelectric hysteresis loops and **b**, **f**, **j** resonance response hysteresis loops under high activation as indicated in the activation maps the inset of **b**, **f**, **j**. These insets have the same colorscales as shown in Fig. 4. Arrows indicate the resonance pathway taken during switching. Phase-field simulations of (top) out-of-plane polarization, (bottom) electrostatic energy of switching under locally applied tip bias when the tip is positioned **c**, **d** within the *c* domain at the valley boundary, **g**, **h** within the *a* domain at the peak boundary, and **k**, **l** within the *a* domain at the valley boundary, wherein the applied tip bias is positive and negative at each position, respectively. Crystallographic orientation is indicated in the bottom left corner. Polarization direction is indicated in the right corner. Color scale for the polarization and electrostatic energy is presented as diverging perceptually uniform colormaps as indicated in the colorbar

controlled manipulation of structure–properties in an automated and self-regulating manner not possible with human-in-the-loop processes.

Methods

Growth of epitaxial PbZr$_{0.2}$Ti$_{0.8}$O$_3$ thin films. Four hundred-nm-thick PbZr$_{0.2}$Ti$_{0.8}$O$_3$ thin films were synthesized using pulsed-laser deposition by ablating a ceramic target of Pb$_{1.1}$Zr$_{0.2}$Ti$_{0.8}$O$_3$ using a KrF excimer laser (248 nm, LPX 305; Coherent), in an on-axis geometry with a 60 mm target-to-substrate spacing. The PbZr$_{0.2}$Ti$_{0.8}$O$_3$ films were grown on 30 nm Ba$_{0.5}$Sr$_{0.5}$RuO$_3$-buffered NdScO$_3$ (110) single-crystal substrates which were affixed to the heater using Ag paint. The Ba$_{0.5}$Sr$_{0.5}$RuO$_3$ bottom electrodes were grown at a heater temperature of 750 °C in a dynamic oxygen pressure of 20 mTorr, by ablating a ceramic Ba$_{0.5}$Sr$_{0.5}$RuO$_3$ target (Praxair) at a laser fluence and a repetition rate of 1.8 J/cm^2 and 2 Hz, respectively. The PbZr$_{0.2}$Ti$_{0.8}$O$_3$ films were grown at a heater temperature of 600 °C in a dynamic oxygen pressure of 50 mTorr, with a laser fluence and repetition frequency of 1.9 J/cm2 and 14 Hz, respectively. Following growth, all heterostructures were cooled to room temperature in a static oxygen pressure of 760 Torr at 5 °C/min.

Band-excitation piezoresponse spectroscopy. BEPS studies were performed at the Center for Nanophase Materials Science (CNMS) at Oak Ridge National Laboratory (ORNL) using a custom Cypher (Asylum Research) atomic force microscope controlled with a Labview- and Matlab-based controller. A bipolar-triangular-switching waveform was applied using a conductive scanning-probe tip in a square grid measuring the cantilever response caused by the band-excitation waveform in the time domain. Following processing with a fast-Fourier transform, the cantilever resonance response was fit to a simple harmonic oscillator model, allowing the extraction of piezoresponse amplitude, phase, cantilever resonance frequency, and dissipation. The use of band excitation for these measurements is crucial as it minimizes effects from changing tip–sample contact resonances that can alter the observed response, enabling consistent measurements of piezoresponse throughout multiple dimensions (that is, frequency, spatial, voltage, time, and so on; Supplementary Note 3). All measurements were carried out using Pt/Ir-coated probe tips (NanoSensor PPP-EFM). Switching spectroscopy measurements were measured at a resonance frequency of ~132 kHz (with a bandwidth of 60 kHz). The DC voltage was chosen such that the piezoelectric hysteresis loops were saturated in both the positive and negative direction. The local piezoresponse was measured at remanence (following a dwell time of 0.5 ms), with a BE waveform of sinc character (peak-to-peak voltage of 1 V).

Neural-network structure and training. The LSTM RNN autoencoders were built in Keras using the Tensorflow backend. The network trained on the piezoresponse data had four encoding and decoding layers each of size 128. Dropout within the encoding and decoding layer was fixed at 20%. The low-dimensional embedding layer had a size of 16 and l_1 regularization ($\lambda = 1 \times 10^{-5}$). Batch-normalization layers were included prior to and following the low-dimensional embedding layer. The network was trained using Adam as an optimizer with an initial learning rate ($L = 3 \times 10^{-5}$) for 16,000 epochs. For the analysis of the resonance data the network used was identical to the network used for the piezoresponse data with the exception that the network was trained for 22,000 epochs. Training was completed using a local workstation equipped with a NVIDIA Titan X graphics processing unit (GPU) or on the Savio supercomputer cluster equipped with GPU nodes with NVIDIA K80 GPUs. To accelerate the training of the generated responses formed, after training the autoencoder sufficiently, the weights through the low-dimensional layer were fixed and the decoder was trained for one million epochs without dropout.

Phase-field simulations. A three-dimensional model was applied to simulate the evolution of ferroelectric polarizations (P_i ($i = 1, 2, 3$)) of the PbZr$_{0.2}$Ti$_{0.8}$O$_3$ (PZT) thin film by numerically solving the time-dependent Landau–Ginzburg–Devonshire (LGD) equations:[64]

$$\frac{\partial P_i(x,t)}{\partial t} = -L \frac{\delta F}{\delta P_i(x,t)}, \; i = 1, \; 2, \; 3 \qquad (1)$$

in which P_i is the polarization vector, x is the spatial position, t is the time, L is the kinetic coefficient related to the domain wall mobility, and F is the total free energy as shown below:[65]

$$F = \int_V \left[f_{Land}(P_i) + f_{Grad}\left(P_{i,j}\right) + f_{Elas}\left(P_i, \varepsilon_{ij}\right) + f_{Elec}(P_i, E_i) \right] dV \qquad (2)$$

in which $f_{Land}(P_i), f_{Grad}\left(P_{i,j}\right), f_{Elas}\left(P_i, \varepsilon_{ij}\right), f_{Elec}(P_i, E_i)$ represent the LGD free energy density, gradient energy density, elastic energy density, and electrostatic energy density, respectively. Details of these energy density terms as well as the coefficients related to these energy terms are collected from literature.[66] Here we adopt a sixth-order polynomial expansion of P_i for $f_{Land}(P_i)$, and choose the dielectric constant to be $\kappa = 50$ for PZT. The gradient energy coefficients are set to

be $G_{11}/G_{110} = 0.6$, where $G_{110} = 1.73 \times 10^{-10}$ C^{-2} m^4 N.[67] The simulation size is a realistic three-dimensional geometry sampled on a fine grid mesh of $128\Delta x \times 128\Delta x \times 32\Delta x$, where the grid size $\Delta x = 1.0$ nm. The film and substrate thickness are $20\Delta x$ and $10\Delta x$, respectively. A semi-implicit spectral method[68] is used to solve the time-dependent LGD equation, with periodic boundary conditions applied in x_1 and x_2 directions, and thin film boundary conditions applied in x_3 direction. The initial structure consists of $(100)_a/(001)_c$ preset domain structure. The entire thin film is subjected to a homogeneous 0.3% tensile strain by the substrate. Electric bias is modeled using a Lorentz function $\varphi(x,y) = \frac{\varphi_0 \gamma^2}{(r-a)^2 + \gamma^2}$ where r is the distance from the tip and γ is the half-width at half-maximum (HWHM) of applied bias $\left(\varphi_0\right)$. The tips are located near the c/a domain boundaries as described in the main text. The average polarization in a $5\Delta x \times 5\Delta x \times 6\Delta x$ cuboid volume near the tip center is collected to calculate the hysteresis loop.

Data availability

References

1. Martin, L. W. & Rappe, A. M. Thin-film ferroelectric materials and their applications. *Nat. Rev. Mater.* **2**, 1–14 (2017). 16087.
2. Schlom, D. G. D. G. et al. Elastic strain engineering of ferroic oxides. *MRS Bull.* **39**, 118–130 (2014).
3. Damodaran, A. R. et al. New modalities of strain-control of ferroelectric thin films. *J. Phys. Condens. Matter* **28**, 1–36 (2016). 263001.
4. Lee, K. S., Choi, J. H., Lee, J. Y. & Baik, S. Domain formation in epitaxial Pb (Zr,Ti)O$_3$ thin films. *J. Appl. Phys.* **90**, 4095–4102 (2001).
5. Ganpule, C. S. et al. Imaging three-dimensional polarization in epitaxial polydomain ferroelectric thin films. *J. Appl. Phys.* **91**, 1477–1481 (2002).
6. Damodaran, A. R. R. et al. Three-state ferroelastic switching and large electromechanical responses in PbTiO3 thin films. *Adv. Mater.* **29**, 1–9 (2017). 1702069.
7. Agar, J. C. et al. Highly mobile ferroelastic domain walls in compositionally graded ferroelectric thin films. *Nat. Mater.* **15**, 549–556 (2016).
8. Agar, J. C. et al. Complex evolution of built-in potential in compositionally-graded PbZr$_{1-x}$Ti$_x$O$_3$ thin films. *ACS Nano* **9**, 7332–7342 (2015).
9. Yadav, A. K. et al. Observation of polar vortices in oxide superlattices. *Nature* **530**, 198–201 (2016).
10. Zubko, P. et al. Negative capacitance in multidomain ferroelectric superlattices. *Nature* **534**, 524–528 (2016).
11. Mundy, J. A. et al. Atomically engineered ferroic layers yield a room-temperature magnetoelectric multiferroic. *Nature* **537**, 523–527 (2016).
12. Daniels, J. E. et al. Neutron diffraction study of the polarization reversal mechanism in [111]c-oriented Pb(Zn$_{1/3}$Nb$_{2/3}$)O$_3$-XPbTiO$_3$. *J. Appl. Phys.* **101**, 1–7 (2007). 104108.
13. Xu, R. et al. Ferroelectric polarization reversal via successive ferroelastic transitions. *Nat. Mater.* **14**, 79–86 (2015).
14. Chen, Z. H., Damodaran, A. R., Xu, R., Lee, S. & Martin, L. W. Effect of asymmetry mismatch on the domain structure of rhombohedral BiFeO3 thin films. *Appl. Phys. Lett.* **104**, 1–5 (2014). 182908.
15. Khan, A. I., Marti, X., Serrao, C., Ramesh, R. & Salahuddin, S. Voltage-controlled ferroelastic switching in Pb(Zr$_{0.2}$Ti$_{0.8}$)O$_3$ thin films. *Nano Lett.* **15**, 2229–2234 (2015).
16. Feigl, L., McGilly, L. J., Sandu, C. S. & Setter, N. Compliant ferroelastic domains in epitaxial Pb(Zr,Ti)O$_3$ thin films. *Appl. Phys. Lett.* **104**, 1–4 (2014). 172904.
17. McGilly, L. J., Yudin, P., Feigl, L., Tagantsev, A. K. & Setter, N. Controlling domain wall motion in ferroelectric thin films. *Nat. Nanotechnol.* **10**, 145–150 (2015).
18. Feigl, L. et al. Controlled creation and displacement of charged domain walls in ferroelectric thin films. *Sci. Rep.* **6**, 1–7 (2016). 31323.
19. Jiang, J. et al. Temporary formation of highly conducting domain walls for non-destructive read-out of ferroelectric domain-wall resistance switching memories. *Nat. Mater.* **17**, 49–56 (2018).
20. Sharma, P. et al. Nonvolatile ferroelectric domain wall memory. *Sci. Adv.* **3**, e1700512 (2017).
21. Nelson, C. T. et al. Domain dynamics during ferroelectric switching. *Science* **334**, 968–971 (2011).

22. Hart, J. L. et al. Electron-beam-induced ferroelectric domain behavior in the transmission electron microscope: toward deterministic domain patterning. *Phys. Rev. B* **94**, 1–7 (2016). 174104.

23. Somnath, S., Belianinov, A., Kalinin, S. V. & Jesse, S. Full information acquisition in piezoresponse force microscopy. *Appl. Phys. Lett.* **107**, 1–4 (2015). 263102.

24. Jesse, S. et al. Band excitation in scanning probe microscopy: recognition and functional imaging. *Annu. Rev. Phys. Chem.* **65**, 519–536 (2014).

25. Ahn, Y. et al. Photoinduced domain pattern transformation in ferroelectric-dielectric superlattices. *Phys. Rev. Lett.* **119**, 1–6 (2017). 57601.

26. Laanait, N. et al. Full-field X-ray reflection microscopy of epitaxial thin-films. *J. Synchrotron Radiat.* **21**, 1252–1261 (2014).

27. Kalinin, S. V., Sumpter, B. G. & Archibald, R. K. Big–deep–smart data in imaging for guiding materials design. *Nat. Mater.* **14**, 973–980 (2015).

28. Perozzi, B., Al-Rfou, R. & Skiena, S. Deepwalk: online learning of social representations. In *Proc. 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 701–710 (New York, 2014).

29. Gkotsis, G. et al. Characterisation of mental health conditions in social media using Informed Deep Learning. *Sci. Rep.* **7**, 1–10 (2017). 45141.

30. Tshitoyan, V. et al. Unsupervised word embeddings capture latent knowledge from materials science literature. *Nature* **571**, 95–98 (2019).

31. Vinyals, O., Toshev, A., Bengio, S. & Erhan, D. Show and tell: a neural image caption generator. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 07–12 June, 3156–3164 (Boston, 2015).

32. dos Santos, C. & Gatti, M. Deep convolutional neural networks for sentiment analysis of short texts. In *Proc. COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, 69–78 (Dublin, 2014).

33. Severyn, A. & Moschitti, A. Twitter sentiment analysis with deep convolutional neural networks. In *Proc. 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 959–962 (2015).

34. Schmidhuber, J. Deep learning in neural networks: an overview. *Neural Netw.* **61**, 85–117 (2015).

35. Goodfellow, I., Bengio, Y., Courville, A. & Bengio, Y. *Deep Learning*. Vol. 1 (MIT Press, Cambridge, 2016).

36. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).

37. Szegedy, C., Ioffe, S., Vanhoucke, V. & Alemi, A. A. Inception-v4, inception-resnet and the impact of residual connections on learning. *AAAI* **4**, 12 (2017).

38. Sutskever, I., Vinyals, O. & Le, Q. V. Sequence to sequence learning with neural networks. In *Advances in Neural Information Processing Systems*, 3104–3112 (Montréal, 2014).

39. Wu, Y., et al. Google's neural machine translation system: bridging the gap between human and machine translation. *arXiv Preprint. arXiv1609.* 08144 (2016).

40. Wu, C. H. & McLarty, J. W. *Neural Networks and Genome Informatics*. Vol. 1 (Elsevier, 2012).

41. Dery, L. M., Nachman, B., Rubbo, F. & Schwartzman, A. Weakly supervised classification in high energy physics. *J. High Energy Phys.* **2017**, 1–5 (2017). 145.

42. Naul, B., Bloom, J. S., Pérez, F. & van der Walt, S. A recurrent neural network for classification of unevenly sampled variable stars. *Nat. Astron.* **2**, 151–155 (2017).

43. Zhang, Y. & Kim, E.-A. Quantum loop topography for machine learning. *Phys. Rev. Lett.* **118**, 1–4 (2017). 216401.

44. Ch'ng, K., Carrasquilla, J., Melko, R. G. & Khatami, E. Machine learning phases of strongly correlated fermions. *Phys. Rev. X* **7**, 1–9 (2017). 31038.

45. Bohrdt, A., et al. Classifying snapshots of the doped Hubbard model with machine learning. *Nat. Phys.* https://doi.org/10.1038/s41567-019-0565-x (2019)

46. Xie, T. & Grossman, J. C. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Phys. Rev. Lett.* **120**, 1–6 (2018). 145301.

47. Segler, M. H. S., Preuss, M. & Waller, M. P. Planning chemical syntheses with deep neural networks and symbolic AI. *Nature* **555**, 604–610 (2018).

48. Xu, W. & LeBeau, J. M. A deep convolutional neural network to analyze position averaged convergent beam electron diffraction patterns. *Ultramicroscopy* **188**, 59–69 (2018).

49. Ziatdinov, M., Maksov, A. & Kalinin, S. V. Learning surface molecular structures via machine vision. *NPJ Comput. Mater.* **3**, 1–9 (2017). 31.

50. Borodinov, N. et al. Deep neural networks for understanding noisy data applied to physical property extraction in scanning probe microscopy. *npj Comput. Mater.* **5**, 1–8 (2019). 25.

51. Ghosh, K. et al. Deep learning spectroscopy: neural networks for molecular excitation spectra. *Adv. Sci.* **6**, 1–7 (2019). 1801367.

52. Zhang, Y. et al. Machine learning in electronic-quantum-matter imaging experiments. *Nature* **570**, 484–490 (2019).

53. Rem, B. S., et al. Identifying quantum phase transitions using artificial neural networks on experimental data. *Nat. Phys.* https://doi.org/10.1038/s41567-019-0554-0 (2019)

54. Neumayer, S. M. et al. Surface chemistry controls anomalous ferroelectric behavior in lithium niobate. *ACS Appl. Mater. Interfaces* **10**, 29153–29160 (2018).

55. Caprioli, R. M., Farmer, T. B. & Gile, J. Molecular imaging of biological samples: localization of peptides and proteins using MALDI-TOF MS. *Anal. Chem.* **69**, 4751–4760 (1997).

56. Neacsu, C. C., Dreyer, J., Behr, N. & Raschke, M. B. Scanning-probe Raman spectroscopy with single-molecule sensitivity. *Phys. Rev. B* **73**, 1–8 (2006). 193406.

57. Kim, Y.-M. et al. Direct observation of ferroelectric field effect and vacancy-controlled screening at the $BiFeO_3/La_xSr_{1-x}MnO_3$ interface. *Nat. Mater.* **13**, 1019–1025 (2014).

58. Cueva, P., Hovden, R., Mundy, J. A., Xin, H. L. & Muller, D. A. Data processing for atomic resolution electron energy loss spectroscopy. *Microsc. Microanal.* **18**, 667–675 (2012).

59. Kluyver, T., et al. in Positioning and Power in Academic Publishing: Players, Agents and Agendas (eds. Loizides, F. & Schmidt, B.) 87–90 (IOS Press, 2016).

60. Young, T., Hazarika, D., Poria, S. & Cambria, E. Recent trends in deep learning based natural language processing [Review Article]. *IEEE Comput. Intell. Mag.* **13**, 55–75 (2018).

61. Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural Comput.* **9**, 1735–1780 (1997).

62. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15**, 1929–1958 (2014).

63. Core Layers. *Keras Documentation*. https://keras.io/layers/core/#dense.

64. Chen, L.-Q. Phase-field method of phase transitions/domain structures in ferroelectric thin films: a review. *J. Am. Ceram. Soc.* **91**, 1835–1844 (2008).

65. Li, Y. L., Hu, S. Y., Liu, Z. K. & Chen, L. Q. Phase-field model of domain structures in ferroelectric thin films. *Appl. Phys. Lett.* **78**, 3878–3880 (2001).

66. Haun, M. J., Zhuang, Z. Q., Furman, E., Jang, S. J. & Cross, L. E. Thermodynamic theory of the lead zirconate-titanate solid solution system, part III: Curie constant and sixth-order polarization interaction dielectric stiffness coefficients. *Ferroelectrics* **99**, 45–54 (1989).

67. Xue, F. et al. Phase field simulations of ferroelectrics domain structures in $PbZr_xTi_{1-x}O_3$ bilayers. *Acta Mater.* **61**, 2909–2918 (2013).

68. Chen, L. Q. & Shen, J. Applications of semi-implicit Fourier-spectral method to phase field equations. *Comput. Phys. Commun.* **108**, 147–158 (1998).

Acknowledgements

Author contributions

J.C.A., L.W.M., R.K.V and S.V.K. designed and conceived the experiments. J.C.A., B.N., J.M., S.v.d.W. and J.B. developed the concept and trained the deep-learning neural network autoencoder. J.C.A., S.P. and R.K.V. conducted the band excitation piezoresponse force microscopy experiments. S.P. and J.C.A. synthesized the films using pulsed-laser deposition. J.C.A., R.Y., Y.C. and L.Q.C. designed and conducted the phase-field simulations. The manuscript was written by J.C.A., L.W.M., R.K.V. and S.V.K. All authors were involved in revising the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41467-019-12750-0.

Correspondence and requests for materials should be addressed to J.C.A. or L.W.M.

Reprints and permission information is available at http://www.nature.com/reprints

Peer Review Information *Nature Communications* thanks Bai-Xiang Xu and other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.